Communications Research Centre

MAXIMUM ENTROPY SPECTRAL ANALYSIS AND RADAR SIGNAL PROCESSING

by

R.W. Herring

This work was sponsored by the Department of National Defence, Research and Development Branch under Project No. 33C06.

DEPARTMENT OF COMMUNICATIONS
MINISTÈRE DES COMMUNICATIONS

CRC REPORT NO. 1330

CANADA

OTTAWA, JANUARY 1980

COMMUNICATIONS RESEARCH CENTRE

DEPARTMENT OF COMMUNICATIONS
CANADA

MAXIMUM ENTROPY SPECTRAL ANALYSIS AND RADAR SIGNAL PROCESSING

by

R.W. Herring

(Radio and Radar Research Branch)

CRC REPORT NO. 1330

OTTAWA

This work was sponsored by the Department of National Defence, Research and Development Branch, under Project No. 33C06.

TABLE OF CONTENTS

MAXIMUM ENTROPY SPECTRAL ANALYSIS AND RADAR SIGNAL PROCESSING

by

R.W. Herring

ABSTRACT

The theory and derivation of the maximum-entropy method of spectral analysis and the Burg algorithm, and the potential applicability of these techniques to radar signal processing, are reviewed. This material is presented in a readily comprehensible form for assessment by the practicing radar engineer. Topics such as computational complexity, statistical properties, adaptive clutter suppression and angular spectrum estimation are discussed. It is concluded that, at least at present, maximum-entropy spectral analysis is not a panacea for radar signal processing, due to its substantial computational requirements, and the relatively unknown and potentially unfavorable statistical properties of data analyzed using the Burg algorithm. A promising area of applicability appears to be that of adaptive Doppler filtering.

1. INTRODUCTION

1.1 DOPPLER AND ANGULAR SPECTRA OF RADAR ECHOES

Radar echoes received from reflectors having a radial component of velocity with respect to the radar have their frequency spectra shifted because of the Doppler effect. Such Doppler shifts can in principle be used to distinguish between wanted targets and unwanted "clutter" echoes, if the characteristics of the motions of the targets and clutter are sufficiently different to render their Doppler spectra distinguishable by the radar signal

processor. Examples of clutter are echoes from mountains, raindrops, the ground, trees, birds, the surface of the ocean, insects, etc.. Sometimes although such clutter targets are not of principal concern, it is desirable to be aware of their presence, such as when flocks of birds suddenly appear near aircraft flight and approach paths.

Radars are restricted in range and angular resolution due to their finite signal bandwidths and non-zero antenna beamwidths. This means that radars have only finite range and angular resolution, so that targets within a "resolution cell" are not resolved but are seen as a single echo. Targets of different types contained in the same resolution cell, however, can have their own characteristic Doppler spectra. Therefore, if it is possible to examine the Doppler spectrum of echoes from within a given cell, knowledge of the targets in that cell can be inferred.

A spectral estimation problem also arises in the case of sampled-aperture radar systems. Radar echoes are received by an array of antennas as a superposition of electromagnetic waves. Each wave varies sinusoidally across the array with a spatial frequency that is proportional to the sine of its angle-of-arrival with respect to the boresight of the array. The angles-of-arrival of the various echoes can be inferred if the spatial frequency spectrum of such data can be estimated.

1.2 THE REQUIREMENT TO ESTIMATE SPECTRA USING AS FEW DATA AS POSSIBLE

For the case of either Doppler or angular spectrum estimation, the number of contiguous data is limited. For the case of angular spectrum estimation, this number is equal to the number of receivers. For the case of Doppler spectrum estimation, the number of contiguous data is limited by the number of pulses used to illuminate a given resolution cell, or "the number of hits on target". For a surveillance radar, the number of hits on target is usually determined by design factors such as how often it is required to search the total surveillance volume, the antenna beamwidth, and the pulse-repetition frequency. For a tracking radar, it might well be possible to increase the number of hits on target in volumes of particular interest, where a wanted target is suspected to exist.

In any of these cases, the spectral characteristics of the data may be changing with time, so that the data are not "stationary". Thus for several reasons it is desirable to be able to estimate spectra using as few data as possible.

1.3 A SUMMARY OF CLASSICAL FOURIER SPECTRAL ANALYSIS AND ITS LIMITATIONS

The classical methods for estimating spectra are based on the Fourier transform. These Fourier techniques are equivalent to least-squares fitting a predetermined set of sine waves to any given set of data. Such data can consist either of samples of amplitude as a function of time, such as are obtained from a coherent radar when a set of successive echoes from a given range are considered, or, of samples of an estimate of the autocorrelation function derived by averaging lagged cross-products of the sampled amplitude data.

In the Fourier analysis of either amplitude or autocorrelation data, the amplitudes and phases of the predetermined set of sine waves are adjusted to minimize the mean-square difference between the data and the fitted sine waves. In practice this fitting operation is simply the discrete Fourier transform.

The resolution and statistical properties of the Fourier transform methods are well known (e.g., [1]). For example, if it is known that a signal to be analyzed consists of two sine waves of different frequencies, then it becomes difficult to estimate the difference between these frequencies when this difference is less than the reciprocal of the length of the data record. That is, if the length of a data record is T seconds, then spectral details having extent in the frequency domain of less than 1/T Hz are obscured. It is also well known that the discontinuities at the ends of the data record cause sidelobes in the Fourier spectrum, an effect known as "spectral leakage", and that spectral leakage can be controlled by multiplying the data with a windowing function before performing the Fourier transform. The penalty exacted by such windowing is a further loss in spectral resolution, so that spectral details of width greater than 1/T are obscured.

Statistically, it is known that if sampled amplitude data comprised of noise with Gaussian statistics are to be analyzed, then estimates of the power spectrum separated in frequency by 1/T are statistically independent and have standard deviations equal to their means. Statistical reliability can be improved either by averaging power spectra derived from non-overlapping sets of data, or by using windowing functions to reduce the spectral resolution.

In summary, then, it can be seen that the spectral resolution capability of the classical Fourier transform method is inherently limited to being approximately the reciprocal of the length of the data record or worse, so that even if it is known on a physical basis that there is information in the spectrum of bandwidth less than this limit, the technique will cause such detail to be obscured. Such is the case for example when target echo classification on the basis of Doppler spectral estimates is attempted by using classical Fourier spectral analysis on data sets that are "too short".

1.4 ALTERNATIVE METHODS FOR SPECTRAL ANALYSIS

Within the past decade, some alternative methods for spectral analysis have found considerable success in the fields of seismic and oil-exploration geophysics, and in speech analysis and transmission. These alternative methods have also aroused interest in the sonar and radar communities.

The alternative methods differ fundamentally from the Fourier methods. The Fourier methods tacitly assume that the data have been generated by a set of independent harmonic generators having frequencies (0/T), (±1/T), (±2/T), etc., where T is the length of the set of data being analyzed. The alternative methods to be discussed here are based on a different tacit assumption. That assumption is: given a set of data, it is possible to design a filter which has as its output random "white" noise when the data is applied to its input. Such a filter is known as a "whitening" filter. An estimate of the spectrum of the data can then be taken to be proportional to the inverse of the power

transfer function of the whitening filter (see Figure 1). The spectral estimation techniques thus consist of techniques for designing whitening filters using only the available data. This approach arises naturally in the course of the formulation of maximum-entropy method.

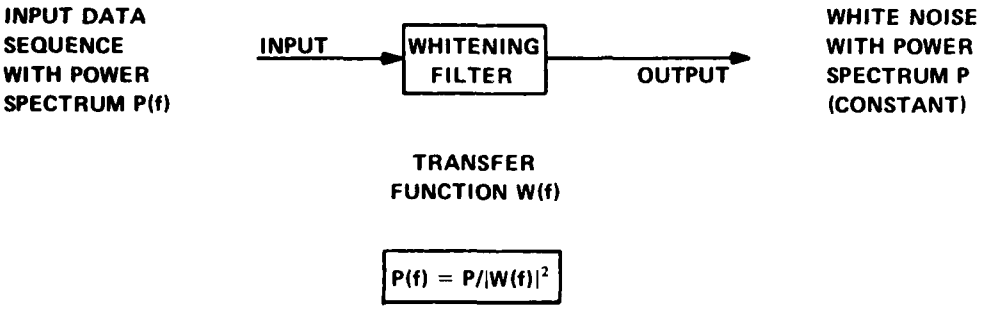


Figure 1. Whitening-Filter Method of Power Spectrum Estimation

For speech analysis, a model for the whitening filter can be based on the known physical characteristics of the human voice. Similarly, for geophysical signal processing, knowledge of the physical processes generating the signals to be analyzed is of aid. For the case of radar signals, it is not yet clear whether or what physical models can be used to provide a priori information regarding the generation of the whitening filters.

1.5 THE MAXIMUM-ENTROPY METHOD

A technique which leads to the generation of a whitening filter and which has helped stimulate current research in the area of radar applications is known as the maximum-entropy method. The details of this method and its theoretical formulation are outlined in Section 2.

It must be noted that the "true" maximum-entropy method is based on the assumption that not time-series amplitude data, such as would be obtained from a sequence of radar echoes from a given range, azimuth and (perhaps) elevation cell are available, but rather that a set of autocorrelation data is available. Such autocorrelation data are not available from present-day radars.

1.6 THE BURG ALGORITHM

To obviate the need for estimating autocorrelation data, the Burg algorithm was developed. The details of this algorithm and its theoretical foundations are reviewed in Section 3. The Burg algorithm generates whitening filters directly from time-series amplitude data, in a relatively efficient manner which reflects the gist of the maximum-entropy technique.

1.7 STATISTICAL BEHAVIOR OF THE MAXIMUM-ENTROPY AND BURG TECHNIQUES

Radar detection is closely related to statistical decision theory. Therefore it is necessary to understand the statistical effects of any procedure applied to the data prior to where any detection decision is made. To date, the statistical behavior of spectra derived using the maximum-entropy method have been studied only asymptotically [2], and the statistical behavior of Burg spectra have been studied only by means of computerized Monte Carlo simulations using simulated data [3,4] or very limited quantities of real data [3,5]. It is on the basis of its statistical behavior that the applicability of the Burg method to radar signal processing will be confirmed or rejected. This topic, which is presently unresolved, is discussed in Section 3.2 and Section 5.

2. THE BASIS AND DERIVATION OF THE MAXIMUM-ENTROPY METHOD OF SPECTRAL ANALYSIS*

The maximum-entropy method of spectral analysis is based on a line of reasoning proposed by J.P. Burg in the 1960s [6,7] and finally published in detail by him in 1975 [8]. He began by supposing that we have available M+1 *error-free* autocorrelation data $R(m\Delta t)$ for a set of M+1 lags: $R(0), R(\Delta t), R(2\Delta t), \ldots, R(M\Delta t)$. He proceeded to show that, instead of simply Fourier transforming this autocorrelation data to obtain a power spectrum, it is possible to devise an alternative procedure based on finding a power spectrum which is both consistent with the given autocorrelation data and also has maximum-entropy in the sense defined by Shannon's Information Theory. The advantage of this technique is supposed to be that it is not subject to the smoothing effects introduced when a truncated series of autocorrelation data are Fourier transformed. This Section outlines the details of the derivation of the maximum-entropy method of spectral analysis and shows how it differs from conventional Fourier techniques. It also lays the foundation for Section 3, where the Burg algorithm for the spectral analysis of amplitude time-series data is described and derived.

2.1 DEFINITION OF THE AUTOCORRELATION FUNCTION

The autocorrelation function of a time-dependent signal $x(t)$ can be defined as

$$R(m\Delta t) = \langle x^*(t)x(t+m\Delta t)\rangle \tag{2.1}$$

where the brackets $\langle\ \rangle$ denote the expected value of the quantity they contain. The asterisk * denotes that if the data corresponding to $x(t)$ are complex (that is, if in-phase and quadrature components are available), then the complex conjugate is to be taken. The time variable t can be either discrete or continuous, but it is assumed that the autocorrelation data are available at

---

* Sections 2 and 3 are based closely on the derivation contained in Chapter II.A of Burg's Ph.D. Thesis [8].

the discrete time-lag intervals $m\Delta t$, where m is a positive or negative integer, or zero.

Notice that the definition of $R(m\Delta t)$ implies that $R(m\Delta t)$ is independent of the actual times at which the signal is observed. This means that the data x(t) are assumed to have statistics that are stationary and do not vary with the passage of time. The supposition that such autocorrelation data are available is usually unrealistic in the radar situation, but in some sonar and geophysical situations, good estimates of the autocorrelation function can be made.

2.2 THE FOURIER-TRANSFORM RELATION BETWEEN THE AUTOCORRELATION FUNCTION AND THE POWER SPECTRUM

In general, a sampled autocorrelation function $R(m\Delta t)$ and its associated power spectrum P(f), where f is frequency, are related by the Fourier transform pair

$$P(f) = \Delta t \sum_{m=-\infty}^{\infty} R(m\Delta t)\exp(-j2\pi fm\Delta t) \tag{2.2}$$

and

$$R(m\Delta t) = \int_{-1/2\Delta t}^{1/2\Delta t} P(f)\exp(j2\pi fm\Delta t)df \tag{2.3}$$

We note that, since power is always real, eqns. (2.2) and (2.3) each imply that

$$R(-m\Delta t) = R^*(m\Delta t) \tag{2.4}$$

This result is also consistent with the definition of $R(m\Delta t)$ as given by eqn. (2.1). We also note that the range of integration is over $\pm 1/2\Delta t$ rather than $\pm\infty$, since the data are sampled.

If we have available only (M+1) autocorrelation data, we could obtain an estimate of the power spectrum by simply truncating the infinite sum of eqn. (2.2) at $m = \pm M$ and arbitrarily setting $R(m\Delta t)$ to zero for $m = \pm(M+1)$ and beyond. But this procedure introduces a step discontinuity into the autocorrelation data, which can lead to problems in the form of physically unrealistic estimates of negative power at some frequencies. The step discontinuities can be thought of as causing a "ringing" in the spectral domain.

Another way to look at this is to consider that the autocorrelation data have been modulated (or multiplied) by a square-wave gate function. This means that the true power spectrum corresponding to the untruncated autocorrelation data has been convolved with the spectrum of the square-wave gating

function, and it is this latter spectrum, which has both positive and negative amplitudes, that introduces the apparent negative powers into the estimated power spectrum.

This problem of negative power estimates can be sidestepped by choosing to multiply the autocorrelation data by some weighting function which decays to zero at $m = \pm M$ and has no negative amplitude components in its frequency spectrum. A typical example is the triangular weighting function, which has its maximum value at m=0 and decays linearly to zero at $m = \pm(M+1)$. Such non-uniform weighting eliminates spurious negative power estimates, but at the expense of smoothing the estimated power spectrum more than does the simple uniform square-wave weighting. There is no way of avoiding this tradeoff between stability and smoothing when using classical Fourier-transform processing.

2.3 THE MAXIMUM-ENTROPY CRITERION

Burg's approach was to try to estimate the spectrum of the process generating the given autocorrelation data without making any a priori assumptions about the unknown data for lags $\pm(M+1)\Delta t$ and beyond. He began by noting that there are an infinite number of power spectra which can be Fourier transformed using eqn. (2.3) to yield autocorrelation functions identical to the given set of sampled data over the range of lags for which the data are available. Beyond that range, of course, all these autocorrelation functions will differ.

The problem then becomes: which one of this infinite set of power spectra to choose? To answer this question, Burg used the criterion of "information entropy" as defined for power spectra by Shannon's Information Theory [9]. A good discussion of the motivation for this choice can be found in Ref. [10]. This entropy H is defined to within a scale factor by the formula

$$H = \Delta t \int_{-1/2\Delta t}^{1/2\Delta t} \log_e[P(f)]df \tag{2.5}$$

Let it be assumed that all the power spectra under consideration are constrained to have the same total power P_o, where

$$P_o = \int_{-1/2\Delta t}^{1/2\Delta t} P(f)df \tag{2.6}$$

Then for

$$P(f) = P_o\Delta t \tag{2.7}$$

that is, P(f) is a flat or a white power spectrum, the entropy H has its maximum value H_o, where

$$H_o = \log_e(P_o \Delta t) \tag{2.8}$$

We also note by comparing eqn. (2.6) with eqn. (2.3) when m=0 that

$$P_o = R(0) \tag{2.9}$$

It is relatively straightforward to convince ourselves that any non-white spectrum will have smaller entropy than a white spectrum having the same total power. We can start by noting that any power spectrum P(f) must have a mean value of $P_o\Delta t$, from eqn. (2.6). Then P(f) can be expressed in the form

$$P(f) = P_o \Delta t \, [1 + p_1(f)] \tag{2.10}$$

so that

$$H = H_o + \Delta t \int_{-1/2\Delta t}^{1/2\Delta t} \log_e[1 + p_1(f)]df \tag{2.11}$$

For the sake of simplicity we will assume that the magnitude of $p_1(f)$ is much less than one, so that

$$|p_1(f)| \ll 1 \tag{2.12}$$

and we can write

$$\log_e[1 + p_1(f)] \approx p_1(f) - \frac{1}{2} p_1^2(f) + \ldots \tag{2.13}$$

Then from eqn. (2.10) we can write

$$H - H_o \approx \Delta t \int_{-1/2\Delta t}^{1/2\Delta t} [p_1(f) - \frac{1}{2} p_1^2(f) + \ldots]df \tag{2.14a}$$

$$\approx -\frac{1}{2} \Delta t \int_{-1/2\Delta t}^{1/2\Delta t} p_1^2(f)df < 0 \tag{2.14b}$$

where we have noted from eqn. (2.10) that the average value of $p_1(f)$ must be zero. Thus we see that $H<H_o$, so that for a given signal power level, a white noise signal indeed has maximum entropy. Although we have only proven this result for a special case, it is true in general.

This, then, is the criterion that Burg used to select one of the infinite sets of possible power spectra: *in the absence of any further information about the process generating the data,* a reasonable choice for

the power spectrum would be one which both satisfied the constraints imposed by the given autocorrelation data and also had the maximum entropy, as defined by eqn. (2.5).

We have already seen that, where there are no constraints due to the autocorrelation function, a white noise spectrum is chosen, since white noise has the greatest possible entropy. White noise is characterized by its random and unpredictable nature. Thus, by inference, choosing a power spectrum which is constrained by the autocorrelation data and yet has maximum entropy in the face of these constraints, suggests that this power spectrum corresponds to the most random possible process consistent with the observed data. Our next task is to find out how to estimate this spectrum.

2.4 THE CONSTRAINED ENTROPY-MAXIMIZATION PROBLEM AND THE FUNCTIONAL FORM OF THE SPECTRAL ESTIMATE

We can estimate the maximum-entropy spectrum by setting up and then solving a constrained maximization problem. This can be done by using Lagrange's undetermined multipliers. We begin by defining the function $\Phi[P(f)]$:

$$\Phi[P(f)] = \Delta t \int_{-1/2\Delta t}^{1/2\Delta t} \log_e P(f)\,df$$

$$+ \sum_{m=-M}^{m} \lambda_m \left[R(m\Delta t) - \int_{-1/2\Delta t}^{1/2\Delta t} P(f) \exp(j2\pi f m\Delta t)\,df \right] \qquad (2.15)$$

where the λ_ms are the undetermined multipliers, and the terms in the square bracket are the constraint conditions derived from eqn. (2.3), which specifies that the Fourier transform of the power spectrum must match the M+1 given values of the autocorrelation function.

If we now take the derivative of $\Phi[P(f)]$ with respect to $P(f)$ and set the result equal to zero, we get

$$\Delta t \int_{-1/2\Delta t}^{1/2\Delta t} \frac{1}{P(f)}\,df - \sum_{m=-M}^{m} \lambda_m \int_{-1/2\Delta t}^{1/2\Delta t} \exp(j2\pi f m\Delta t)\,df = 0 \qquad (2.16)$$

Solving eqn. (2.16) for $P(f)$ yields

$$P(f) = \Delta t \Big/ \sum_{m=-M}^{M} \lambda_m \exp(j2\pi f m\Delta t) \qquad (2.17)$$

and we now have the functional form of the maximum entropy estimate of P(f). But we have yet to determine the λ_ms in terms of the autocorrelation data.

We already know that if P(f) is a power spectral density, then P(f) must be real and positive. This means that we must have

$$\lambda_{-m} = \lambda_m^* \tag{2.18}$$

so that the imaginary terms in the denominator of eqn. (2.17) cancel. Then we can write

$$\sum_{m=-M}^{M} \lambda_m \exp(j2\pi fm\Delta t) = P_M^{-1} \left| 1 - \sum_{m=1}^{M} \alpha(m,M)\exp(-j2\pi fm\Delta t) \right|^2 \tag{2.19}$$

where the M+1 independent unknown λ_ms have been replaced by the real unknown P_M and the M complex unknown $\alpha(m,M)$s. It will turn out that an explicit solution for the λ_ms is not required; it is only important that we can write P(f) as

$$P(f) = \frac{P_M \Delta t}{\left| 1 - \sum_{m=1}^{M} \alpha(m,M)\exp(-j2\pi fm\Delta t) \right|^2} \tag{2.20}$$

2.5 DETERMINATION OF THE COEFFICIENTS IN THE SPECTRAL ESTIMATE FROM THE AUTOCORRELATION DATA

Now we are faced with the problem of how to determine the $\alpha(m,M)$s from the $R(m\Delta t)$s. We can do this by going back to eqn. (2.3) and evaluating the integral using P(f) as given by eqn. (2.20). We can do this most easily by changing variables as follows. We let

$$z = \exp(j2\pi f\Delta t) \tag{2.21}$$

so that

$$df = \frac{z^{-1}dz}{j2\pi\Delta t} \tag{2.22}$$

We note that for $f = \pm 1/2\Delta t$, the corresponding values for z is z=-1 in both cases. Thus the integral of eqn. (2.3) has been transformed into a contour integral around the unit circle in the complex z-plane, and can be written as

$$R(m\Delta t) = \frac{P_M}{j2\pi} \oint \frac{z^{m-1}dz}{\left[- \sum_{k=0}^{M} \alpha(k,M)z^{-k} \right] \left[- \sum_{k=0}^{M} \alpha^*(k,M)z^{k} \right]} \tag{2.23}$$

where we have defined

$$\alpha(0,M) = -1 \tag{2.24}$$

to simplify the notation. Next, using eqn. (2.23), we form the summation

$$-\sum_{n=o}^{M} \alpha(n,M)R[(r-n)\Delta t] = \frac{P_M}{j2\pi}\oint \frac{z^{r-1}\left[-\sum_{n=0}^{M} \alpha(n,M)z^{-n}\right] dz}{\left[-\sum_{k=0}^{M} \alpha^*(k,M)z^{k}\right]\left[-\sum_{k=0}^{M} \alpha(k,M)z^{-k}\right]} \tag{2.25a}$$

$$= \frac{P_M}{j2\pi}\oint \frac{z^{r-1}\, dz}{\left[-\sum_{k=0}^{M} \alpha^*(k,M)z^{k}\right]} \tag{2.25b}$$

We note without proof that it is always possible to formulate the polynomial in the denominator of eqn. (2.25b) so that it does not have any zeroes inside the unit circle. Then for $r \geq 1$, the integral of eqn. (2.25b) is zero, by Cauchy's integral theorem. For the case r=0, we observe that there is a single pole at z=0. For the case of a single pole, Cauchy's integral theorem can be stated as

$$\frac{1}{j2\pi}\oint \frac{f(z)}{z}\, dz = f(0) \tag{2.26}$$

so that the right-hand side of eqn. (25b) is equal to P_M, since $f(0) = -\alpha(0,M) = -(-1) = 1$.

Finally we have the system of M+1 complex equations in M+1 unknowns required to be solved to determine the M complex $\alpha(m,M)$s and the real unknown P_M:

$$R(r) - \sum_{m=1}^{M} \alpha(m,M)R(r-m) = P_M \quad (r=0) \tag{2.27a}$$

$$R(r) - \sum_{m=1}^{M} \alpha(m,M)R(r-m) = 0 \quad (r=1,2,3\ldots,M) \tag{2.27b}$$

(From now on we will often set $\Delta t=1$ to simplify the notation.)

2.6 THE MAXIMUM-ENTROPY SPECTRAL ESTIMATE AND PREDICTION-ERROR FILTERS

The set of coefficients $\{1,-\alpha(1,M),-\alpha(2,M),\ldots,-\alpha(M,M)\}$ in eqns. (2.27a) and (2.27b) define what is called a prediction-error filter (or PEF) of order M and length (M+1). This prediction-error filter, as the name suggests, uses

the M values of the autocorrelation function, R(-M) to R(-1), to predict an estimated value for R(0); then it subtracts this estimate from the actual value of R(0) and outputs the difference, P_M. In this manner the PEF attempts to account for all the predictable behavior of the autocorrelation function that can be expressed in terms of a weighted sum of M of its terms. P_M is the power that cannot be accounted for on the basis of this model.

The eqns. (2.27b) can be used to extrapolate the autocorrelation function beyond the limit r=M. This extrapolated autocorrelation function is the Fourier transform of the maximum-entropy power spectrum, as required by Fourier transform relation of eqn. (2.3). This extrapolation of the autocorrelation function is optimum only in the sense that it satisfies the constrained entropy maximization criterion. The usual radar signal processing approach of matched filtering has nothing whatever in common with this result.

Examination of eqn. (2.20) shows that the maximum-entropy power spectrum is inversely proportional to the squared magnitude of the Fourier transform of the prediction-error filter coefficients; that is, if we write

$$F(z,M) = 1 - \sum_{m=1}^{M} \alpha(m,M) z^{-1} \tag{2.28}$$

where

$$z = \exp(j2\pi f \Delta t) \tag{2.21}$$

then we can rewrite eqn. (2.20) for the maximum-entropy power spectrum as

$$P(f) = P_M \Delta t / |F(z,M)|^2 \tag{2.29}$$

This means that the PEF can be thought of as a whitening filter, for if we apply a signal which has a power spectrum P(f) to the input terminals of such a PEF, then the power spectrum of the output signal would be $P_M \Delta t$, which is a constant, independent of frequency. Constant power spectra correspond to white noise; hence the name whitening filter.

2.7 SOLVING FOR THE COEFFICIENTS OF THE MAXIMUM-ENTROPY PREDICTION-ERROR FILTER

The set of eqns. (2.27a) and (2.27b) can be written in matrix form:

$$\begin{bmatrix} R(0) & R(-1) & R(-2) & \cdots & R(-M) \\ R(1) & R(0) & R(-1) & & \vdots \\ R(2) & R(1) & R(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & R(-1) \\ R(M) & \cdots & \cdots & R(1) & R(0) \end{bmatrix} \begin{bmatrix} 1 \\ -\alpha(1,M) \\ -\alpha(2,M) \\ \vdots \\ -\alpha(M,M) \end{bmatrix} = \begin{bmatrix} P_M \\ 0 \\ 0 \\ \\ 0 \end{bmatrix} \qquad (2.30)$$

It can be seen that the matrix of autocorrelation data has a highly ordered structure. This structure, in which the elements of each diagonal are identical is called Toeplitz by mathematicians. In such an (M+1)x(M+1) Toeplitz matrix there are $(M+1)^2$ entries, only 2M+1 of which are distinct. For the matrix of eqn. (2.30), because of the conjugate-symmetry relation of eqn. (2.4), there are only M+1 independent elements (i.e., the matrix is also Hermitian).

Usually it is necessary to solve a matrix equation such as eqn. (2.30) by inverting the matrix. When the matrix is Toeplitz, however, such equations can be solved by a relatively simple procedure called the Levinson recursion algorithm.

We begin by using only the term R(0). Then the PEF has only one term, and eqn. (2.30) becomes

$$R(0)x1 = P_0 \qquad (2.31)$$

Next we shall see how, as we take each value R(1),R(2),...,R(M) in turn, we can uniquely extend the length of the PEF by one coefficient each time, in a highly efficient manner. We start the derivation of the recursion algorithm by assuming that we have already solved the matrix equation

$$\begin{bmatrix} R(0) & R(-1) & \cdots & R(-K) \\ R(1) & R(0) & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ R(K) & \cdots & \cdots & R(0) \end{bmatrix} \begin{bmatrix} 1 \\ -\alpha(1,K) \\ \vdots \\ -\alpha(K,K) \end{bmatrix} = \begin{bmatrix} P_K \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (2.32)$$

for some value of K, K=0,1,2,...,M-1, and that we have obtained the solution for the PEF of order K and length K+1. We observe that we can rewrite eqns. (2.27a) and (2.27b), by taking their complex conjugate and using eqn. (2.4), as

$$R(-r) - \sum_{m=1}^{M} \alpha^*(m,M)R(m-r) = P_M \quad (r=0) \tag{2.33a}$$

$$R(-r) - \sum_{m=1}^{M} \alpha^*(m,M)R(m-r) = 0 \quad (r=1,2,\ldots,M) \tag{2.33b}$$

Equations (2.33a) and (2.33b) can also be written in matrix form:

$$\begin{bmatrix} R(0) & R(-1) & R(-2) & \cdots & R(-M) \\ R(1) & R(0) & R(-1) & & \vdots \\ R(2) & R(1) & R(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & R(0) & R(-1) \\ R(M) & \cdots & \cdots & R(1) & R(0) \end{bmatrix} \begin{bmatrix} -\alpha^*(M,M) \\ -\alpha^*(M-1,M) \\ -\alpha^*(M-2,M) \\ \vdots \\ -\alpha^*(1,M) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ P_M \end{bmatrix} \tag{2.34}$$

If we compare eqns. (2.34) and (2.30), we can see that the same results are obtained when the sequence of PEF coefficients is reversed and the coefficients are conjugated. This observation is the key to the Levinson recursion.

We want the output of each successively higher order PEF to have the same functional form; that is, an output sequence with one non-zero term and with all other terms being zeroes. Therefore, using eqns. (2.32) and (2.34), we write

$$\begin{bmatrix} R(0) & R(-1) & \cdots & R(-K) & R(-K-1) \\ R(1) & R(0) & & R(-K+1) & R(-K) \\ \vdots & & \ddots & & \vdots \\ R(K) & R(K-1) & \cdots & R(0) & R(-1) \\ R(K+1) & R(K) & \cdots & R(1) & R(0) \end{bmatrix} \left(\begin{bmatrix} 1 \\ -\alpha(1,K) \\ \vdots \\ -\alpha(K,K) \\ 0 \end{bmatrix} - \alpha(K+1,K+1) \begin{bmatrix} 0 \\ -\alpha^*(K,K) \\ \vdots \\ -\alpha^*(1,K) \\ 1 \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} P_K \\ 0 \\ \vdots \\ 0 \\ Q_K \end{bmatrix} - \alpha(K+1,K+1) \begin{bmatrix} Q_K^* \\ 0 \\ \vdots \\ 0 \\ P_K \end{bmatrix} \right) \tag{2.35a}$$

$$= \begin{bmatrix} P_{K+1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \qquad (2.35b)$$

Comparing eqns. (2.35a) and (2.35b) with eqn. (2.32), we see that we have introduced one new piece of data, R(K+1), and three new parameters: $\alpha(K+1,K+1)$, P_{K+1} and Q_K. From eqn. (2.35a) we can write

$$Q_K = R(K+1) - \sum_{k=1}^{K} \alpha(k,K)R(K+1-k) \qquad (2.36)$$

Comparison of eqn. (2.36) and eqn. (2.27b) for M=K and r=k+1 shows that Q_K is the difference between the extrapolated estimate of R(K+1) produced by the PEF of order K and the actual value of R(K+1).

If Q_K is zero, then we can see from the right-hand sides of eqns. (2.35a) and (2.35b) that $\alpha(K+1,K+1)$ must also be zero, and the value of R(K+1) and all subsequent values of the autocorrelation function are perfectly predictable from the PEF we already have. Also, we can see from eqn. (2.40) below that $P_K = P_{K+1} = P_{K+2} = \ldots$ becomes constant. This special case arises when the process generating the signal whose autocorrelation function data are being analyzed can be modelled as the output of a feedback network containing exactly K feedback loops and being excited by an input signal consisting of uncorrelated white noise with power density P_K. Such a process is often referred to as being "autoregressive" or "all-pole", and therefore maximum-entropy spectral analysis turns out to be based on an autoregressive model of the signal-generating process. Of course, if the signal-generating process cannot be so modelled, then Q_K vanishes only in the limit $K \to \infty$. There is further discussion of this topic in Sections 3.7 and 5.2.1.

If Q_K is non-zero, then by equating the right-hand sides of eqns. (2.35a) and (2.35b) we can write

$$P_K - \alpha(K+1,K+1)Q_K^* = P_{K+1} \qquad (2.37)$$

and

$$Q_K - \alpha(K+1,K+1)P_K = 0 \qquad (2.38)$$

From eqn. (2.38) we can write

$$\alpha(K+1,K+1) = Q_K/P_K \qquad (2.39)$$

and by eliminating Q_K between eqns. (2.37) and (2.38) we can write

$$P_{K+1} = (1- |\alpha(K+1,K+1)|^2)P_K \tag{2.40}$$

Equation (2.40) has the interesting implication that either

$$|\alpha(K+1,K+1)| \leq 1 \tag{2.41}$$

or else we could have negative power at the output of the PEF. This possibility does not arise if we have proper autocorrelation data, for as we stated following eqn. (2.25b), it is always possible to choose a PEF which has all its roots inside the unit circle on the complex z-plane. There is a theorem in algebra which states that the coefficient of the last term of the z-transform of the PEF as we have formulated it (i.e., $\alpha(K+1,K+1)$) is equal to (± 1) times the product of its K+1 roots. If all the roots have magnitude less than one, then it follows that inequality (2.41) will be obeyed.

As an aside, we note there is one special case where equality can be achieved for eqn. (2.41), and that is when all K roots of the PEF lie on the unit circle. This special case corresponds to having a signal comprised of K complex sinusoidal terms and no additive noise. Such a signal is not truly autoregressive in nature, for, although it can be modelled by the output of a feedback network which has K feedback loops, no input signal is required. The output signal is determined strictly by the initial conditions when the network was activated.

Also, we can see from eqn. (2.40) that if $|\alpha(K+1,K+1)| = 1$, then $P_{K+1} = 0$. Going back to eqn. (2.38), we see that then we must also have $Q_{K+1} = 0$, so that the Levinson recursion terminates for a PEF of order K+1 when $|\alpha(K+1,K+1)| = 1$.

Returning now to the main argument of the derivation, it remains to be shown how to derive the remaining coefficients $\alpha(1,K+1),\alpha(2,K+1),\ldots,\alpha(K,K+1)$ of the $K+1^{th}$ order PEF from the K^{th} order PEF. We do this by noting that from the left-hand side of eqn. (2.35a) we can write

$$\begin{bmatrix} 1 \\ -\alpha(1,K) \\ -\alpha(2,K) \\ \vdots \\ -\alpha(K,K) \\ 0 \end{bmatrix} - \alpha(K+1,K+1) \begin{bmatrix} 0 \\ -\alpha^*(K,K) \\ -\alpha^*(K-1,K) \\ \vdots \\ -\alpha^*(1,K) \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -\alpha(1,K+1) \\ -\alpha(2,K+1) \\ \vdots \\ -\alpha(K,K+1) \\ -\alpha(K+1,K+1) \end{bmatrix} \tag{2.42}$$

or, in general

$$\alpha(k,K+1) = \alpha(k,K) - \alpha(K+1,K+1)\alpha^*(K+1-k,K) \tag{2.43}$$

$$k = 1,2,\ldots,K$$

Thus we can see that the sequence of PEFs from order zero to order M are completely specified if the set of coefficients

$$P_o,\alpha(1,1),\alpha(2,2),\ldots,\alpha(K,K),\ldots,\alpha(M,M) \tag{2.44}$$

are known. For historic reasons, these $\alpha(K,K)$s are known as reflection coefficients, because of their mathematical analogy to the reflection coefficients which arise in the theory of wave propagation through a layered medium. It is interesting to note that this set of coefficients is comprised of one real-valued term and M complex-valued terms, the same as the original set of M+1 autocorrelation values.

Finally, we recapitualte the Levinson recursion scheme. We begin at K=0 and by letting $P_o=R(0)$. Then we apply eqns. (2.36), (2.39), (2.43), and (2.40) for K=0,1,2,...,M to obtain a sequence of PEFs. We end by applying eqn. (2.29) to obtain the M^{th}-order maximum entropy power spectral estimate. (Note that for K=0, the summation in eqn. (2.36) is ignored, since the upper limit of the summation index is less than the lower limit.)

2.8 EXTRAPOLATING THE AUTOCORRELATION FUNCTION

Sometimes it is desired to extrapolate the autocorrelation function data beyond the M+1 values assumed available. We can see by recalling the discussion of Section 2.3 and by examining eqn. (2.27b) that such an extrapolation arises naturally in the maximum-entropy formulation. By removing the restriction $r \leq M$ on eqn. (2.27b) we obtain the formula for the maximum-entropy extrapolation $\hat{R}(r)$ of the autocorrelation data:

$$\hat{R}(r) = \sum_{m=1}^{M} \alpha(m,M)\hat{R}(r-m) \quad (r = M+1,M+2,\ldots,\infty) \tag{2.45}$$

$\hat{R}(r-m)$ is replaced by $R(r-m)$ in the right-hand side of eqn. (2.44) for $r-m \leq M$; i.e., the given data are used wherever available.

2.9 INVERTING THE AUTOCORRELATION MATRIX

Sometimes it is necessary to compute the inverse of the autocorrelation matrix. The Levinson recursion can be used to do this efficiently, and we shall show how the inverse of either the matrix of given autocorrelation data or the inverse of the maximum-entropy extrapolation of the autocorrelation matrix can be calculated.

Equations (2.30) and (2.45) can be combined to give

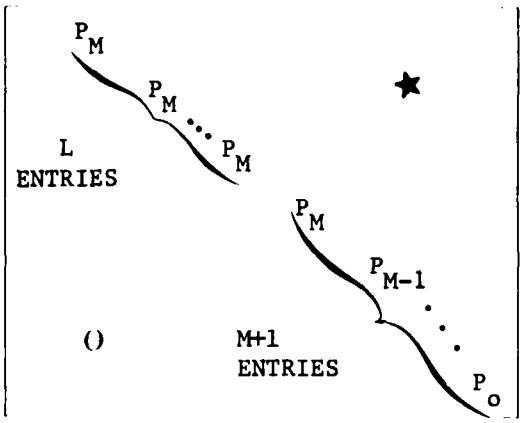
$$\begin{bmatrix} R(0) & R(-1) & \cdots & R(-M) & \hat{R}(-M-1) & \cdots & \hat{R}(-M-L) \\ R(1) & R(0) & & R(-M) & & & \\ \vdots & & \ddots & & \ddots & & \hat{R}(-M-1) \\ R(M) & & & & & & R(-M) \\ \hat{R}(M+1) & R(M) & & & & & \\ \vdots & \vdots & & & & R(0) & R(-1) \\ \hat{R}(M+L) & \cdots & \hat{R}(M+1) & R(M) & \cdots & R(1) & R(0) \end{bmatrix} \begin{bmatrix} 1 \\ -\alpha(1,M) \\ \\ \\ -\alpha(M,M) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} P_M \\ 0 \\ \vdots \\ \\ \\ \\ \\ 0 \end{bmatrix} \tag{2.46}$$

where we have assumed that L values of the extrapolated autocorrelation function have been included and therefore that L zeroes have been appended to augment the length of the PEF. Of course, if L=0, then eqn. (2.46) reduces to eqn. (2.30).

In order to simplify the notation, we shall now denote the (M+1+L) x (M+1+L) autocorrelation matrix by [R(M+L)]. We want now to determine its inverse, $[R(M+L)]^{-1}$. This we can do by noting that

$$[R(M+L)] \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ -\alpha(1,M) & 1 & & & & & & & & \\ -\alpha(2,M) & -\alpha(1,M) & \ddots & & & & & & & \\ \vdots & \vdots & & 1 & & & & & & \\ -\alpha(M,M) & -\alpha(M-1,M) & -\alpha(1,M) & & 1 & & & & & \\ 0 & -\alpha(M,M) & & & -\alpha(1,M-1) & \ddots & & & & \\ & 0 & & & & & & & & \\ & & & -\alpha(M-1,M) & & & & & 1 & 0 \\ 0 & 0 & \cdots & -\alpha(M,M) & -\alpha(M-1,M-1) & \cdots & & -\alpha(1,1) & & 1 \end{bmatrix}$$

(M+1+L)x(M+1+L)



$$= \qquad (2.47)$$

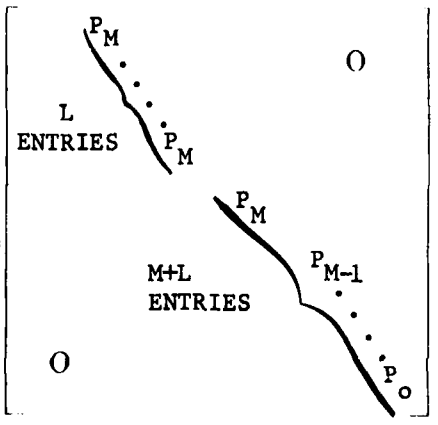
$(M+1+L) \times (M+1+L)$

The $(M+1+L) \times (M+1+L)$ matrix of PEFs and augmented PEFs on the left-hand side of eqn. (2.47) we shall denote by $[\alpha(M+L)]$. The explicit contents above the diagonal of the matrix with the Ps along its main diagonal, denoted by the large asterisk, are not required in what follows below.

If we now let $[\alpha(M+L)]^+$ denote the conjugate transpose of $[\alpha(M+L)]$ we can write

$$[\alpha(M+L)]^+[R(M+L)][\alpha(M+L)] = [P(M+L)] \qquad (2.48)$$

where



$$[P(M+L)] = \qquad (2.49)$$

The reason that $[P(M+L)]$ is diagonal is that both $[\alpha(M+L)]^+$ and the right-hand side of eqn. (2.47) are upper-triangular matrices, and therefore their product must also be upper-triangular. However, the left-hand side of eqn. (2.48) is Hermitian, so that the upper-triangular product matrix must also be Hermitian, which can only be true if the off-diagonal elements are zeroes.

The inverse of the diagonal matrix [P(M+L)] we can write by inspection:

$$[P(M+L)]^{-1} = \begin{bmatrix} P_M^{-1} & & & & & & \\ & \ddots & & & & O & \\ & & P_M^{-1} & & & & \\ & & & \ddots & & & \\ & & & & P_M^{-1} & & \\ & & & & & \ddots & \\ & O & & & & & P_o^{-1} \end{bmatrix} \tag{2.50}$$

We now find that a few simple manipulations of eqn. (2.48) will lead us to the desired result. For brevity we will drop the explicit dependence on the argument (M+L). Post-multiplying eqn. (2.48) by $[P]^{-1}$ yields

$$[\alpha]^{+}[R][\alpha][P]^{-1} = [I] \tag{2.51}$$

where [I] denotes [I(M+L)], the (M+1+L) x (M+1+L) identity matrix. Pre-multiplying both sides of eqn. (2.51) by [α] and post-multiplying by $[\alpha]^{+}$ yields

$$[\alpha][\alpha]^{+}[R][\alpha][P]^{-1}[\alpha]^{+} = [\alpha][\alpha]^{+} \tag{2.52}$$

Finally, pre-multiplying both sides of eqn. (2.52), first by $[\alpha]^{-1}$ and then by $[\alpha]^{+-1}$ (since [α] is clearly non-singular) yields the desired result:

$$[R][\alpha][P]^{-1}[\alpha]^{+} = [I] \tag{2.53}$$

or

$$[R(M+L)]^{-1} = [\alpha(M+L)][P(M+L)]^{-1}[\alpha(M+L)]^{+} \tag{2.54}$$

Thus we see that, given a set of M reflection coefficients and P_o, the zero-lag value of the autocorrelation function, the inverse of the maximum-entropy extrapolation of the autocorrelation matrix can be computed. Further examination of the inverse of the Toeplitz autocorrelation matrix would show that this inverse matrix is both persymmetric (i.e., symmetric about the diagonal from its lower-left to upper-right corner) and Hermitian, which means that the (M+1+L) x (M+1+L) matrix has at most (M+1+L)(M+3+L)/4 independent elements if (M+1+L) is even, and $(M+2+L)^2/4$ independent elements if (M+1+L) is odd. This observation is useful in simplifying the implementation of eqn. (2.54) to derive the actual elements of the inverse autocorrelation matrix.

For the particular case L=0 (i.e., the autocorrelation matrix is not extrapolated before inversion), it is straightforward to derive the equations for the elements of the inverse autocorrelation matrix [7]. If we let $\tilde{R}(i,j)$ denote an element of this matrix, then

$$\tilde{R}(i,j) = \tilde{R}^*(j,i) = \tilde{R}(M-i,M-j) = \tilde{R}^*(M-j,M-i)$$

$$= \sum_{k=0}^{n} \alpha(M-k,i-k)\alpha^*(M-k,j-k)/P_{M-k} \qquad (2.55)$$

$$i,j = 0,1,\ldots,M$$

where n is the smaller of the indices i,j and we let $\alpha(M-k,0) = -1$ as in eqn. (2.24).

2.10 DERIVING THE AUTOCORRELATION MATRIX FROM A GIVEN SET OF REFLECTION COEFFICIENTS

In Section 3 et seq. we shall derive the Burg algorithm, which allows us to compute a set of reflection coefficients and an estimate of P_o directly from a set of amplitude time-series data. We might then want to estimate the autocorrelation function from the estimated reflection coefficients. Therefore it is useful to derive the equations for computing the autocorrelation function from a given set of reflection coefficients.

We can easily obtain the required relations from eqns. (2.27a) and (2.27b). From eqn. (2.27a) we have

$$R(0) = P_o \qquad (2.56)$$

and from eqn. (2.27b) we have

$$R(K) = \sum_{k=1}^{K} \alpha(k,K)R(K-k) \qquad (2.57)$$

$$K = 1,2,\ldots,M$$

where the $\alpha(k,K)$s are derived from the set of reflection coefficients by using eqn. (2.43).

2.11 SUMMARY

We have shown how, given a set of perfect autocorrelation data, a power spectrum which has maximum entropy in the sense of Shannon's Information Theory can be estimated. We have shown that this power spectrum is closely related to a unique set of prediction-error filters which can be computed from the data, and we have derived the algorithm for computing these prediction-error filters. We have also shown how to extrapolate the autocorrelation data using the maximum-entropy criterion, how to invert the extrapolated or unextrapolated autocorrelation matrix, and how to reconstitute the autocorrelation data if a set of reflection coefficients and the value of the zero-log autocorrelation is given instead.

We have not considered the effects of statistical fluctuations in the autocorrelation data, which will exist in any set of measured autocorrelation

data. This rather complicated problem has been considered by Baggeroer [2] who found that, in general, a maximum-entropy spectral estimate based on a given data set will have both a more "spikey" spectrum and a greater variability than a Fourier-transform spectral estimate based on the same data.

3. THE BURG ALGORITHM FOR DETERMINING AUTOREGRESSIVE SPECTRAL ESTIMATES FROM THE TIME-SERIES DATA

We saw in Section 2 how, given a set of autocorrelation data, we could find an estimation of the power spectrum which had maximum entropy. It turned out that the maximum-entropy spectral estimate was inversely proportional to the frequency response of a prediction-error filter (PEF) which was derivable from the autocorrelation data.

In the radar situation, we are often faced with having available only relatively short sets of amplitude time-series data. Sometimes we want to extract the maximum amount of spectral information from these data, and do not want to be subject to the effects of spectral broadening introduced when conventional Fourier-transform techniques are used.

The Burg algorithm for spectral analysis is based on the concept of deriving a set of reflection coefficients and hence a set of PEFs directly from the amplitude time-series data. The logic behind the various steps of the algorithm is prinicpally their analogy with the steps taken when the autocorrelation of the data is given.

We being by assuming that we are given a set of N data $x(n)$: $n=0,1,2,\ldots,N-1$. We want to compute a set of PEFs from these data, and by analogy we want this set of PEFs to have the same properties as if they had been derived from autocorrelation data. In fact, having computed such a set of PEFs, we can then obtain from them a set of autocorrelation estimates by inverting the Levinson recursion scheme as shown in Section 2.9.

These autocorrelation estimates $\tilde{R}_B(m)$ would not be the same as those derived by taking the conventional lagged cross-product average

$$\tilde{R}_c(m) = S(m) \sum_{n=0}^{N-m} x^*(n)x(n+m) \tag{3.1}$$

where $S(m)$ is a scaling factor. (Compare eqns. (3.1) and (2.1).) For reference in what follows, we note that if the scaling factor

$$S(m) = 1/N \tag{3.2}$$

is chosen, that $\tilde{R}_c(m)$ is biased, since then

$$\langle \hat{R}_c(m) \rangle = \left(\frac{N-m}{N}\right) R(m) \tag{3.3}$$

and the expected value of $R_c(m)$ is linearly tapered to zero at $m=N$ lags. If the scaling factor

$$S(m) = 1/(N-m) \tag{3.4}$$

is chosen, then $\hat{R}_c(m)$ is unbiased, since then

$$\langle \hat{R}_c(m) \rangle = \left(\frac{N-m}{N-m}\right) R(m) \tag{3.5}$$

but the expected variance of $\hat{R}_c(m)$ for successively greater values of m becomes greater, due to the smaller number of terms in each successive summation.

3.1 THE RELATION BETWEEN MAXIMUM-ENTROPY AND A MINIMUM POWER OR MINIMUM-MEAN-SQUARE CRITERION

In order to determine a criterion which can be used to estimate a PEF from the time-series data, it is useful to examine the properties of the maximum-entropy PEFs derived from known autocorrelation data. For compactness of notation we let the M^{th}-order PEF be denoted by the column vector $\underline{\alpha}(M)$ where

$$\underline{\alpha}(M) = \begin{vmatrix} 1 \\ -\alpha(1,M) \\ . \\ . \\ . \\ -\alpha(M,M) \end{vmatrix} \tag{3.6}$$

and the conjugate transpose of $\underline{\alpha}(M)$ by $\underline{\alpha}^+(M)$ where

$$\underline{\alpha}^+(M) = [1, -\alpha^*(1,M), \ldots, -\alpha^*(M,M)] \tag{3.7}$$

We also let the (M+1) x (M+1) autocorrelation matrix be denoted by [R(M)] and note that it is equal to its own conjugate transpose.

Now we note that

$$\underline{\alpha}^+(M)[R(M)] = [P_M, 0, \ldots, 0] \tag{3.8a}$$

$$[R(M)]\,\underline{\alpha}\,(M) = [P_M, 0, \ldots, 0]^+ \tag{3.8b}$$

and

$$\underline{\alpha}^+(M)[R(M)]\underline{\alpha}(M) = P_M \tag{3.9}$$

where P_M is the output power as before. We want to know what the output power would be if any other PEF, say $\underline{\beta}(M)$, where

$$\underline{\beta}(M) = \begin{bmatrix} 1 \\ -\beta(1,M) \\ \cdot \\ \cdot \\ \cdot \\ -\beta(M,M) \end{bmatrix} \tag{3.10}$$

were to be used instead. Therefore we consider the expression

$$(\underline{\beta}^+(M)-\underline{\alpha}^+(M))[R(M)](\underline{\beta}(M)-\underline{\alpha}(m)) = \underline{\beta}^+(M)[R(M)]\underline{\beta}(M) - \underline{\alpha}^+(M)[R(M)]\underline{\beta}(M) \tag{3.11a}$$

$$- \underline{\beta}^+(M)[R(M)]\underline{\alpha}(M) + \underline{\alpha}^+(M)[R(M)]\underline{\alpha}(M)$$

$$= \underline{\beta}^+(M)[R(M)]\underline{\beta}(M) - P_M - P_M + P_M \tag{3.11b}$$

$$= \underline{\beta}^+(M)[R(M)]\underline{\beta}(M) - P_M \tag{3.11c}$$

Therefore we find that

$$\underline{\beta}^+(M)[R(M)]\underline{\beta}(M) = P_M + (\underline{\beta}^+(M)-\underline{\alpha}^+(M))[R(M)](\underline{\beta}(M)-\underline{\alpha}(M)) \tag{3.12}$$

We note that we can write, for *any* column vector $\underline{\lambda}(M)$ of length M+1,

$$\underline{\gamma}^+(M)[R(M)]\underline{\gamma}(M) = \sum_{k=0}^{M}\sum_{\ell=0}^{M} \gamma^*(k)R(k,\ell)\gamma(\ell) \tag{3.13}$$

where, by comparison with, say, eqn. (2.32)

$$R(k,\ell) = R(k-\ell) \tag{3.14}$$

Also, from eqn. (2.1) we can write

$$R(k-\ell) = \langle x^*(t)x(t+k-\ell)\rangle \tag{3.15a}$$

or

$$R(k-\ell) = \langle x^*(t-k)x(t-\ell)\rangle \tag{3.15b}$$

or

$$R(k-\ell) = \langle x^*(t+\ell)x(t+k)\rangle \tag{3.15c}$$

if the process generating x(t) is stationary. Then we can write, using eqn. (3.15b),

$$\underline{\gamma}^+(M)[R(M)]\underline{\gamma}(M) = \sum_{k=0}^{M}\sum_{\ell=0}^{M} \gamma^*(k)\langle x^*(t-k)x(t-\ell)\rangle\gamma(\ell) \tag{3.16a}$$

$$= \left\langle \sum_{k=0}^{M} \gamma^*(k)x^*(t-k) \sum_{\ell=0}^{M} \gamma(\ell)x(t-\ell) \right\rangle \tag{3.16b}$$

$$= < \left| \sum_{\ell=0}^{M} \gamma(\ell)x(t-\ell) \right|^2 > \qquad (3.16c)$$

$$\geq 0 \qquad (3.16d)$$

Applying this general result to eqn. (3.12) we can see that any PEF other than the one generated from the autocorrelation data by the Levinson recursion will have an output power greater than P_M. From this Burg inferred that a minimum mean-square-error criterion is appropriate for use in estimating the Burg PEF.

3.2 JUSTIFICATION FOR SIMULTANEOUSLY MINIMIZING THE FORWARD AND BACKWARD OUTPUT POWER FROM PEF

Going back to eqn. (3.16c), we note that we can write

$$\underline{\alpha}^+(M)[R(M)]\underline{\alpha}(M) = < \left| x(t) - \sum_{m=1}^{M} \alpha(m,M)x(t-m) \right|^2 > \qquad (3.17)$$

$$= P_M$$

However, in eqn. (3.1 a) we could just as easily have used eqn. (3.15c) for R(M), to have gotten instead the result

$$\underline{\alpha}^+(M)[R(M)]\underline{\alpha}(M) = < \left| x(t) - \sum_{m=1}^{M} \alpha^*(m,M)x(t+m) \right|^2 > \qquad (3.18)$$

$$= P_M$$

Comparing eqns. (3.17) and (3.18), we see that in the former the PEF is applied to the time series data in the usual forward direction, and that the expected mean-square output is equal to P_M; in the latter equation, the complex conjugate of the PEF is applied to the data in the reverse direction, and again the expected mean-square output is equal to P_M. From this observation Burg inferred that one ought to apply the Burg PEF to the data in both directions simultaneously, and then minimize the mean of the forward and backward output powers.

3.3 COMPUTATION OF THE BURG PEFs

The zero-order PEF is, by definition, the constant "one". Thus the output energy of the zero-order PEF is estimated by

$$E_o = \sum_{n=0}^{N-1} |x(n)|^2 \qquad (3.19)$$

so that E_o/N is an unbiased estimate of P_o.

If the first-order Burg PEF is applied to the data in the forward direction, without going beyond the ends of the data-set, we get a sequence of N-1 outputs, denoted by $f_1(n)$, where

$$f_1(n) = x(n+1) - \beta(1,1)x(n) \quad n=0,1,\ldots,N-2 \tag{3.20}$$

Here $\beta(1,1)$ is the Burg estimate of the first-order reflection coefficient. Similarly, if we apply the conjugate first-order PEF to the data in the backward direction, we get the output sequence

$$b_1(n) = x(n) - \beta^*(1,1)x(n+1) \quad n=0,1,\ldots,N-2 \tag{3.21}$$

Then the output energy of the first order PEF is estimated by

$$E_1 = \frac{1}{2}\sum_{n=0}^{N-2} (|f_1(n)|^2 + |b_1(n)|^2) \tag{3.22}$$

Following the line of reasoning given above, that the maximum-entropy PEF has minimum output energy, we want to estimate $\beta(1,1)$ by minimizing E_1 with respect to $\beta(1,1)$. We do this by expanding eqn. (3.22), taking the first derivative with respect to $\beta(1,1)$, and equating the result to zero. The result is

$$\beta(1,1) = \frac{\sum_{n=0}^{N-2} x^*(n)x(n+1)}{\frac{1}{2}\sum_{n=0}^{N-2} (|x(n)|^2 + |x(n+1)|^2)} \tag{3.23}$$

We notice that the numerator of eqn. (3.23) looks very much like the lagged cross-product average of eqn. (3.1) for a lag of one unit (M=1) and S(m)=1, and the denominator looks something like the zero-lag cross-produce of eqn. (3.1), except that the terms $|x(0)|^2$ and $|x(N-1)|^2$ appear only once; the other N-2 terms appear twice. We also notice that in general

$$|\beta(1,1)| \leq 1 \tag{3.24}$$

which is a consequence of the well-known Schwartz Inequality.

If we take separately the expected values of the denominator and numerator of eqn. (3.23), we note that

$$\frac{<\sum_{n=0}^{N-2} x^*(n)x(n+1)>}{<\frac{1}{2}\sum_{n=0}^{N-2} (|x(n)|^2 + |x(n+1)|^2)>} = \frac{(N-1)R(1)}{\frac{1}{2}(N-1)2R(0)} \tag{3.25a}$$

$$= R(1)/R(0) \tag{3.25b}$$

$$= \alpha(1,1) \tag{3.25c}$$

In deriving eqn. (3.25a) we have used eqn. (2.1), and we have used eqns. (2.36) and (2.38) with K=0 in going from eqn. (3.25b) to eqn. (3.25c). Thus we can see that the first-order Burg reflection coefficient approximates the true maximum-entropy reflection coefficient if the expected values are taken in the manner shown. In general, however, this procedure is unjustifiable, and $\langle\beta(1,1)\rangle \neq \alpha(1,1)$.

We progress now to the estimation of the second-order Burg reflection coefficient. Again we apply the Burg PEF to the data in both the forward and backward directions, without going beyond the ends of the data-set. Then we estimate the mean of the forward and backward output energies, and minimize this energy by adjusting *only* the second-order reflection coefficient.

Applying the second-order Burg PEF in the forward direction, we obtain a sequence of N-2 outputs, denoted by $f_2(n)$, where

$$f_2(n) = x(n+2) - \beta(1,2)x(n+1) - \beta(2,2)x(n)$$

$$n=0,1,\ldots,N-3 \qquad (3.26)$$

Similarly, applying the conjugate second-order Burg PEF in the backward direction, we obtain a sequence of N-2 outputs, denoted by $b_2(n)$, where

$$b_2(n) = x(n) - \beta^*(1,2)x(n+1) - \beta^*(2,2)x(n+2)$$

$$n=0,1,\ldots,N-3 \qquad (3.27)$$

We notice (for example) that the first (n=0) term of each of the sequences $f_2(n)$ and $b_2(n)$ is dependent on the first (n=0), second (n=1) and third (n=2) terms of the original data sequence, and that there are two PEF parameters, $\beta(1,2)$ and $\beta(2,2)$, to be determined.

We recall now that we want to make the Burg PEF as analogous as possible to the maximum entropy PEF. If this is to be the case, then we can use eqn. (2.43) to express $\beta(1,2)$ in terms of $\beta(1,1)$ and $\beta(2,2)$:

$$\beta(1,2) = \beta(1,1) - \beta(2,2)\beta^*(1,1) \qquad (3.28)$$

On a physical basis, it can be considered that $\beta(1,1)$ contains all the information possible concerning how much of the signal can be predicted on the basis by considering the data two at a time. Therefore any additional predictability of the signal which can be discovered on the basis of examining the data three at a time, must be expressable as a function of $\beta(2,2)$ only. However, we also note that this restriction can mean that the absolute minimum output energy achievable if both $\beta(1,2)$ and $\beta(2,2)$ were simultaneously adjusted may not be realized. This is the price paid for forcing a Toeplitz structure on the estimated autocorrelation matrix. In general, sample autocorrelation matrices derived from equi-spaced sampled data are Hermitian, but are not Toeplitz [11].

By substituting eqn. (3.28) into eqn. (3.26) and using eqns. (3.20) and (3.21), we find that $f_2(n)$ is expressable in terms of $f_1(n)$ and $b_1(n)$:

$$f_2(n) = x(n+2) - [\beta(1,1) - \beta(2,2)\beta^*(1,1)]x(n+1) - \beta(2,2)x(n) \quad (3.29a)$$

$$= [x(n+2) - \beta(1,1)x(n+1)] - \beta(2,2)[x(n) - \beta^*(1,1)x(n+1)] \quad (3.29b)$$

$$= f_1(n+1) - \beta(2,2)b_1(n) \quad (3.29c)$$

$$n=0,1,\ldots,N-3$$

Similarly, we find

$$b_2(n) = x(n) - [\beta^*(1,1) - \beta^*(2,2)\beta(1,1)]x(n+1) - \beta^*(2,2)x(n+2) \quad (3.30a)$$

$$= [x(n) - \beta^*(1,1)x(n+1)] - \beta^*(2,2)[x(n+2) - \beta(1,1)x(n+1)] \quad (3.30b)$$

$$= b_1(n) - \beta^*(2,2)f_1(n+1) \quad (3.30c)$$

$$n=0,1,\ldots,N-3$$

For the second-order Burg PEF, the output energy E_2 is given by

$$E_2 = \frac{1}{2}\sum_{n=0}^{N-3}(|f_2(n)|^2 + |b_2(n)|^2) \quad (3.31)$$

Following the same procedure that led to eqn. (3.23), we find that $\beta(2,2)$ is given by

$$\beta(2,2) = \frac{\sum_{n=0}^{N-3} b_1^*(n)f_1(n+1)}{\frac{1}{2}\sum_{n=0}^{N-3}(|b_1(n)|^2 + |f_1(n+1)|^2)} \quad (3.32)$$

Comparing eqns. (3.32) and (3.23) we note the structural similarity. In fact, if we let

$$b_o(n) = x(n) = f_o(n) \qquad n=0,1,\ldots,N-1 \quad (3.33)$$

we can rewrite eqn. (3.23) as

$$\beta(1,1) = \frac{\sum_{n=0}^{N-2} b_o^*(n)f_o(n+1)}{\frac{1}{2}\sum_{n=0}^{N-2}(|b_o(n)|^2 + |f_o(n+1)|^2)} \quad (3.34)$$

In general, we find that

$$\beta(K,K) = \frac{\sum_{n=0}^{N-1-K} b^*_{K-1}(n) f_{K-1}(n+1)}{\frac{1}{2}\sum_{n=0}^{N-1-K} (|b_{K-1}(n)|^2 + |f_{K-1}(n+1)|^2)} \qquad (3.35)$$

$$K=1,2,\ldots,N-1$$

and

$$f_K(n) = f_{K-1}(n+1) - \beta(K,K) b_{K-1}(n) \qquad (3.36)$$

$$b_K(n) = b_{K-1}(n) - \beta^*(K,K) f_{K-1}(n+1) \qquad (3.37)$$

$$n=0,1,\ldots,N-1-K$$

Equation (3.35) tells us that successively higher orders of the Burg PEF reflection coefficients can be calculated by successively computing the single-lag cross-correlation between the forward and backward output signals from the immediately lower order PEF and normalizing by the mean energy of those two output signals, excepting that, in calculating this mean, the last backward output and the first forward output are omitted. Applying Schwartz's inequality to eqn. (3.35) again tells us that, in general,

$$|\beta(K,K)| \leq 1 \qquad (3.38)$$

which conforms with the restriction (2.41) on $|\alpha(K,K)|$ for the MEM estimate.

Finally, then, we have derived the K^{th} order Burg spectral estimate

$$\hat{P}_K(f) = \frac{\Pi_K}{\left|1 - \sum_{k=1}^{K} \beta(k,K)\exp(-j2\pi kf)\right|^2} \qquad (3.39)$$

where, by analogy with eqn. (2.40), the output powers Π_K are defined as

$$\Pi_0 = E_0/N \qquad K=0 \qquad (3.40)$$

and

$$\Pi_K = \Pi_{K-1}\,(1 - |\beta(K,K)|^2) \qquad (3.41)$$

$$K=1,2,3,\ldots,N-1$$

and by analogy with eqns. (2.43) and (3.28)

$$\beta(k,K) = \beta(k,K-1) - \beta(K,K)\,\beta^*(K-k,K-1) \qquad (3.42)$$

$$k=1,2,\ldots,K-1$$
$$K=1,2,3,\ldots,N-1$$

3.4 BIAS OF THE BURG ESTIMATE OF THE AUTOCORRELATION FUNCTION

We saw from eqns. (3.25a-c) that the Burg estimate of the first-order reflection coefficient is roughly equivalent to the true maximum-entropy estimate. Unfortunately, however, the Burg estimate is not in general unbiased.

The existence of bias in the Burg algorithm can be demonstrated by considering the following special case [11]. From eqn. (2.57), for K=1 we can write

$$R(1) = \alpha(1,1)P_o \tag{3.43}$$

where

$$P_o = R(0) \tag{3.44}$$

By analogy, we can also write

$$\hat{R}_B(1) = \beta(1,1)\Pi_o \tag{3.45}$$

where $\hat{R}_B(1)$ can be taken as the Burg estimate of the first lag of the autocorrelation function. If we substitute eqns. (3.23) and (3.19) for N=3 into eqn. (3.45) then we get

$$\hat{R}_B(1) = \frac{x^*(0)x(1) + x^*(1)x(2)}{\frac{1}{2}\left[|x(0)|^2 + 2|x(1)|^2 + |x(2)|^2\right]} \times \frac{|x(0)|^2 + |x(1)|^2 + |x(2)|^2}{3} \tag{3.46}$$

We can see that $\langle\hat{R}_B(1)\rangle$ depends on more than just $\langle x^*(0)x(1)\rangle = \langle x^*(1)x(2)\rangle = R(1)$; it depends on the joint probability distribution of $(x(0),x(1))$.

Following the example [11], we let the data be real, and we let

$$x(0) = u \tag{3.47a}$$

$$x(1) = \frac{1}{\sqrt{2}}(u+v) \tag{3.47b}$$

and

$$x(2) = v \tag{3.47c}$$

where u and v are independent, zero-mean, unit-variance random variables with Gaussian distribution. Then we have

$$\langle x(0)x(1)\rangle = \langle x(1)x(2)\rangle = \frac{1}{\sqrt{2}} \tag{3.48a}$$

and

$$\langle x(0)x(2)\rangle = 0 \tag{3.48b}$$

By substituting eqns. (3.47a-c) into eqn. (3.46), we obtain

$$\hat{R}_B(1) = \frac{1}{\sqrt{2}}\frac{1}{6}\frac{(u+v)^2(3u^2+2uv+3v^2)}{u^2+uv+v^2} \tag{3.49}$$

so that

$$\langle\hat{R}_B(1)\rangle = \frac{1}{\sqrt{2}}\frac{1}{6}\frac{1}{2\pi}\iint_{\infty}^{\infty} dudv\ \exp\left(-\frac{u^2+v^2}{2}\right)\frac{(u+v)^2(3u^2+2uv+3v^2)}{u^2+uv+v^2} \tag{3.50a}$$

$$= \frac{1}{\sqrt{2}}\frac{1}{6}\frac{1}{2\pi}\int_{0}^{\infty} dr\ r^3\ \exp(-r^2/2)\int_{0}^{2\pi} d\theta\ \frac{(C-S)^2(3C^2+2CS+3S^2)}{C^2+CS+S^2} \tag{3.50b}$$

where we have changed to polar coordinates and let $C = \cos\theta$ and $S = \sin\theta$. The integral over r in eqn. (3.50b) is equal to 2 and the integral over θ is equal to $4\pi(2-1/\sqrt{3})$ from [11]. Therefore

$$\langle\hat{R}(1)_B\rangle = \frac{1}{\sqrt{2}}\frac{12-2\sqrt{3}}{9} = \frac{1}{\sqrt{2}} \times 0.9484 \tag{3.51}$$

which is not equal to

$$R(1) = \langle x(0)x(1)\rangle = \langle x(1)x(2)\rangle = \frac{1}{\sqrt{2}} \tag{3.52}$$

Thus we see that the Burg estimate of the first lag of the autocorrelation is biased; but, more important, we see that the bias is inherently dependent on both the specific statistical distribution describing the data and the spectral characteristics of the data as described by the true autocorrelation function. It is this phenomenon that makes theoretical studies of the statistical behavior of the Burg algorithm extremely difficult.

3.5 ESTIMATING THE APPROPRIATE ORDER OF THE BURG SPECTRAL ESTIMATOR

We saw in Section 2.7, in the discussion following eqn. (2.36), that the parameter Q_K is the arithmetic difference between successive maximum-entropy estimates of the autocorrelation functions and the given autocorrelation data when such autocorrelation data are known and available. In particular, we saw that Q_K, the difference between the extrapolated value of the $K+1^{th}$ autocorrelation sample and the actual value of the $K+1^{th}$ autocorrelation sample, goes to zero if the process generating the data can be perfectly modelled by a feedback network containing exactly K feedback loops, being excited by an input signal consisting of white noise. Thus, for this very special case, we have a perfect indicator for determining the correct order of PEF required to model the process.

We also noted in Section 2.7 that a consequence of having $Q_K=0$ was that the $K+1^{th}$ reflection coefficient became zero, so that $P_{K+1} = P_K$ and the out-

put power of the PEF decreased no further as attempts to derive higher-order PEFs were made. This observation might lead us to attempt to use some similar criterion to determine the most appropriate order of PEF with which to model the signal-generating process. However, it is clear that, even if the signal-generating process were truly autoregressive and of order M, in general the statistical fluctuations which normally occur in finite-length data records would be sufficient to prevent $\beta(M+1,M+1)$, the $(M+1)^{th}$ Burg reflection coefficient, from vanishing.

There have been three criteria proposed for estimating the appropriate order M of the Burg PEF used to make a spectral estimate from a given set of amplitude time-series data. We will not attempt a detailed derivation of any of them here, but simply reference their sources.

The first, and most commonly used criterion, is known as Akaike's "final prediction error" (FPE) criterion [12], and is defined as

$$FPE(M) = \Pi_M \left\{ \frac{N+M}{N-M} \right\} \tag{3.53}$$

where Π_M is defined by eqns. (3.41) and (3.42), N is the number of time-series data, and M is the order of PEF currently under consideration. The term M is replaced by (M+1) in both the numerator and the denominator if the mean has been estimated and subtracted from the data. Heuristically, we can note that Π_M is a monotonically decreasing function of M, reflecting the fact that, in general, the more parameters we fit to the data, the better the fit achieved. Conversely, the factor in the braces is a monotonically increasing function of M, reflecting the fact that the more parameters fitted to the data, the less certain is the statistical reliability of each successively higher-order PEF. Statistical fluctuations can arise causing local minima in the FPE estimates from a given set of data; it is usual to choose that value for M which gives the global minimum for the FPE. The FPE was derived as a measure of the mean-square error to be expected when a PEF derived from one set of time-series data is applied to an independent set of time-series data obtained from the same (assumed stationary) source.

A second criterion, also due to Akaike, is called "an information-theoretic criterion" (AIC) [13] and is given in Ref. [14] as

$$AIC(M) = \log_e \Pi_M + 2M/N \tag{3.54}$$

Again the value of M corresponding to the global minimum of AIC(M) is to be chosen. The AIC is a measure of the negative of the log-likelihood ratio of the estimate Π_M of the power of the white-noise excitation of the assumed-autoregressive signal-generating process, expressed as a function of the order of the fitted PEF. The minimum AIC occurs on the average for the most likely to be appropriate value for M, the order of the assumed-autoregressive generating process.

Parzen [15] has defined what he calls a "criterion autoregressive transfer function" (CAT) as a measure of the mean-square error between the "true" PEF of unknown, possible infinite length, and the estimated PEF. This can be done without explicit knowledge of the "true" PEF, and the CAT(M) is defined as

$$CAT(M) = \frac{1}{N} \sum_{m=1}^{M} \frac{N-m}{N\Pi_m} - \frac{N-M}{N\Pi_m} \tag{3.55}$$

Again, the value of M corresponding to the global minimum of CAT(M) is to be chosen.

These criteria are discussed more fully in Ref. [14]; in particular, simulation results are presented which indicate that for many applications, the simplest criterion, the FPE, is the most satisfactory.

4. A BRIEF SUMMARY FOR OTHER TECHNIQUES FOR DETERMINING AUTOREGRESSIVE SPECTRAL ESTIMATES FROM TIME-SERIES DATA

There are several other techniques for making autoregressive spectral estimates from time-series data, but none of them are equivalent to the maximum-entropy method in the same sense as the Burg algorithm. These other methods can be divided into two classes. One class is typified by the Yule-Walker technique, wherein biased or unbiased estimates of the autocorrelation function are first made from the time-series data by means of eqn. (3.1), and then a PEF is estimated by solving the same form of matrix equation as is solved for the maximum-entropy PEF when the autocorrelation data are known exactly. The other class is typified as an unconstrained minimization procedure [16] wherein for each order of PEF, the output power is minimized with respect to all of the PEF coefficients simultaneously (except for the leading term).

4.1 THE YULE-WALKER TECHNIQUE

The Yule-Walker technique exacts one of two penalties in its use, depending on whether biased or unbiased estimates of the autocorrelation are used. If unbiased estimates are used, then loss of spectral resolution can be expected because of the windowing effect described in Section 1.3. If biased estimates are used, then it is possible that a PEF that does not have all its zeroes inside the unit circle (see Section 2.7) can be designed. In this case the extrapolated autocorrelation function derived using eqn. (2.45) would diverge rather than converge to zero as the lag tended to infinity. In terms of effects on the estimated spectrum, it is not clear if there are any that are significant.

An apparent disadvantage of the Yule-Walker techniques is that they require prior computation of autocorrelation estimates before they are executed. Of course if there is an excess of data, it may be possible to compact the data by computing the autocorrelation cross-product sums of eqn. (3.1) as the data are being collected, so that the apparent disadvantage becomes in fact an advantage. It seems unlikely that this would ever be the situation in a radar application.

4.2 THE UNCONSTRAINED MINIMIZATION TECHNIQUE

As mentioned previously, the unconstrained minimization technique requires that for each order of PEF desired, the minimization procedure must be reapplied to the amplitude time-series data. This means that the recursive efficiency of the Burg algorithm is lost. The output power which is minimized can be either the forward-output power, where the filter is applied to the data in the forward direction, or the reverse output power, where the conjugate filter is applied in the reverse direction; or some combination of these powers, such as the average. It has been reported on the basis of simulation studies [16] that minimizing the average power appears to give more accurate spectral estimates, as opposed to minimizing either the forward or reverse power only.

A major computational disadvantage of the unconstrained minimization techniques is that they require the inversion of KxK matrix, where K is the order of the PEF being estimated. The matrix has some symmetry properties which can be used to simplify the inversion process [17], but the matrix does not have Toeplitz symmetry so that the relatively straightforward Levinson recursion technique is not applicable.

5. APPLICABILITY OF THE BURG ALGORITHM TO RADAR SIGNAL PROCESSING

In the time-domain processing of radar-echo data there appear to be three potential applications for the Burg algorithm: clutter spectral estimation and classification, Doppler-shifted target detection and classification, and inverse clutter-correlation matrix estimation. In the spatial-domain processing of radar data from antenna arrays, there appears to be potential application for enhancing angular resolution somewhat beyond that given by conventional beamforming algorithms. As well, estimation of the inverse of the aperture-correlation matrix is possible. This last topic, which is relevant to sidelobe cancellation, is not discussed further here.

In order to assess the feasibility and performance of the algorithm for these applications, the questions of complexity of implementation, statistical properties and biases must be considered. Since the first question is the most tractable, it will be examined first.

5.1 COMPUTATIONAL COMPLEXITY OF THE BURG ALGORITHM

5.1.1 Numbers of Arithmetic Operations Required to Compute the PEF

From eqn. (3.35) it is apparent that when the (N-K+1) $f_{K-1}(n)$s and $b_{K-1}(n)$s are available, then (N-K) complex multiplications are required to compute the numerator of $\beta(K,K)$ and 3(N-K) real additions and 4(N-K) real multiplications are required to compute the denominator. Since a complex multiplication comprises 2 real additions and 4 real multiplications, then we can see that it requires 5(N-K) real additions, 8(N-K)+1 real multiplications and 2 real divisions to evaluate eqn. (3.35)

For K>1, it is also necessary to evaluate eqns. (3.36) and (3.37), which *each* require (N-K) complex additions and (N-K) complex multiplications or 4(N-K) real additions and 4(N-K) real multiplications. Therefore, to compute the sequence of reflection coefficients $\beta(K,K)$, K=1,2,...,M necessary to compute the M^{th}-order Burg PEF requires {N[13M-8]-M[13(M+1)/2-8]} real additions, {N[16M-2]-M[8M-1]} real multiplications and 2M real divisions.

In order to compute spectra from the sequence of M reflection coefficients, it is necessary to compute from them the M^{th}-order PEF by means of eqn. (2.42). It is straightforward to show that it requires 2M(M-1) each of real additions and real multiplications to compute this PEF.

After the PEF has been evaluated, it is possible to obtain unscaled spectral estimates by arbitrarily setting Π_K=1 for K=M in eqn. (3.39). However, if it is required that the spectral power be accurately represented in the sense that

$$\int_{-0.5}^{0.5} P_M(f)df = \Pi_0 \tag{5.1}$$

then it is necessary that Π_M be computed from Π_0 at the cost of 2M real additions and 3M real multiplications. As well, an additional (2N-1) real additions, (2N+1) real multiplications and one real division are required to evaluate Π_0.

In total, then it requires at most [N(13M-6)-3M(3M-1)/2-1] real additions, [16NM-2M(3M-1)+1] real multiplications and (2M+1) real divisions to evaluate all the parameters required to specify the M^{th}-order Burg spectral estimator. Some gains in efficiency can be realized by eliminating redundancies in computing the denominator of eqn. (3.35) [18]; taking these improvements into account reduces the numbers of computations to {[N(9M-2)-M(5M-17)/2]-10} real additions, [4N(3M+1)-M(4M-9)-8] real multiplications, and (2M+1) real divisions. In practice, it would turn out that many of these operations could be done in parallel. The numbers of operations required for representative values of N and M are given in Tables 5.1 and 5.2.

5.1.2 Dynamic Range Considerations

After the PEF of order M and the scale constant Π_M have been evaluated, it is necessary to evaluate eqn. (3.39), which is a function of the continuous variable f, the normalized frequency. Since the spectrum is a true power spectral density estimate, it is possible for there to occur very high spikes of very narrow spectral width, corresponding to a large fraction of the total signal power, whenever there is a highly coherent (e.g., sinusoidal) component present with high SNR in the data. This implies that if narrow-band signals are being sought, then "closely-spaced" evaluations of the power density estimate must be made if such spectral lines are not to be missed.

If we again examine the denominator of eqn. (3.39) we note that it has the form of a discrete Fourier transform, and that peaks in the spectral estimate correspond to minima in this Fourier transform. Since we are most interested in the smallest values of this denominator, it appears that we may well be faced with a dynamic range problem in these regions.

TABLE 5.1

Number of Real Additions Required to Derive an M^{th}-order Burg PEF from N Complex Data

N \ M	1	2	4	8	16	32
8	52	125	256	–	–	–
16	108	253	528	1018	–	–
32	220	509	1072	2138	4030	–
64	444	1021	2160	4378	8574	16006
128	892	2045	4336	8858	17662	34310
256	1788	4093	8188	17818	35838	70918
512	3580	8189	17392	35738	72190	144134

TABLE 5.2

Number of Real Multiplications Required to Derive an M^{th}-order Burg PEF from N Complex Data

N \ M	1	2	4	8	16	32
8	125	218	380	–	–	–
16	253	442	796	1408	–	–
32	509	890	1628	3008	5384	–
64	1021	1786	3292	6208	11656	21016
128	2045	3578	6620	12608	24200	45848
256	4093	7172	13276	25408	49288	95512
512	8189	14330	26588	51008	99464	194840

Estimates of the dynamic ranges encountered in the application of the Burg algorithm can be made by examining plots of spectra obtained from data known to contain signals of narrow bandwidth (e.g., [4,10,25]). It appears that dynamic ranges of 40 to 60 dB are common. Although there appear to have been no studies on the magnitudes of the signals internal to an implementation of the Burg algorithm, it is possible to speculate on their expected magnitudes.

Whenever the input data are highly correlated, such as when narrow-band clutter or strong target echoes are present, it is reasonable to expect that the magnitudes of the reflection coefficients will be close to unity and consequently that the magnitudes of the Π_Ks, $f_K(n)$s and $b_K(n)$s will also decrease fairly rapidly as a function of K. As well, for highly correlated inputs with intrinsically narrow spectra, it is to be expected tha the zeroes of the PEF which correspond to the narrow spectral lines will be close to the unit circle in the complex Z-plane. Such zeroes produce very small minima in the denominator of eqn. (2.39).

Thus it seems that there are potential dynamic range problems inherent in the evaluation of all of eqns. (3.35), (3.36), (3.37) and (3.39). Although further investigations are required in order to clarify this issue, it appears that, for fixed-point implementations, word lengths greater than those required for satisfactory operation of conventional finite-length impulse-response (FIR) or DFT Doppler processors may be required, and that the use of floating-point hardware, perhaps with word-lengths comparable to those found in large-scale computers (e.g., 32 to 38 bits) might be necessary.

5.2 STATISTICAL PROPERTIES OF THE BURG ALGORITHM

5.2.1 Stability and Variability

Because of its nonlinear formulation, direct investigation of the statistical properties of the Burg algorithm is extremely difficult. For this reason theoretical studies are usually confined to consideration of the effects of random errors in estimates of the sampled autocorrelation function when spectra are estimated using the methodology of Section 2 [2,19].

Authors who claim superiority of resolution for maximum-entropy methods over the classical Fourier methods in the analysis of time-series amplitude data usually do so on the basis of comparing the effective number of auto-correlation samples (i.e., the order of the PEF plus one) required to success-fully estimate the spectrum of the process under investigation, as compared with the number of autocorrelation samples required by the classical Fourier methods to achieve similar performance (e.g., [2,3,19,20,21]). However, with the exception of References [3] and [21], the number of raw data contributing to the spectral estimates being presented as reliable is very much larger than the orders of the PEFs derived from them.

A typical example is the paper by Moorcroft [20] which considers the Doppler spectra of radar echoes from radio aurora. Due to its design, the radar system produced data in the form of three contiguous but mutually incoherent blocks of 27 time-series samples each. A modified version of the Burg algorithm, which took into account the lack of coherence between the three data blocks, was used to estimate seventh-order PEFs from sets of 81 data. It was found that, on the average, where typical averages were taken over 25 to 150 spectral estimates, the performance of this modified Burg algorithm was indeed superior to that of classical Fourier spectral estimates averaged in the same manner. This was due to two factors. The first was the systematic limitation on the length of the sets of data which could be processed coherently (i.e., Fourier transformed), which limited the inherent resolution of the Fourier spectra. The second was the fact that apparently the physical process generating the Doppler spectra could be adequately modelled by an all-pole whitening-filter model, as discussed in Section 1.4. However, it must be noted that the total number of raw data contributing to a single averaged spectrum could hardly be considered as limited, and the author demonstrated that valid spectral estimates could be derived only by averaging.

Similar studies have also been undertaken using computer-simulated clutter data [3,4] and limited real data [3] which indicate that reliable clutter spectral estimates may perhaps be derived using record lengths as

short as 16 data and averaging over about 10 independent sets of data (e.g., successive scans of a surveillance radar) obtained from a single radar resolution cell.

Doppler signature analyses of jet-engined aircraft have been made from the processing single blocks of real (i.e., single-channel as opposed to quadrature-detected) data of length N=64 [22]. These results were obtained using an experimental short-range highly stable scanning radar, and the data analyzed were from a range cell known to contain a target.

5.2.2 Statistical Distribution of Burg Spectral Estimates [4]

The only work currently (May 1979) known to this author concerning estimates of the statistical distribution of Burg spectral estimates is that of Chan and Haykin [4]. These investigators have performed extensive computer simulation studies of the response of the Burg algorithm to input data consisting of computer-generated white random noise of various statistical distributions, colored random noise (to simulate radar clutter data) derived by filtering white gaussian noise data, and white gaussian noise plus a D.C. line (corresponding to very narrow stationary clutter), in the presence and absence of sinusoidal signals of fixed frequencies and powers (corresponding to non-fluctuating Doppler-shifted targets). The salient observations and conclusions are summarized below; copies of the complete report are available from its authors.

5.2.2.1 Statistical Distribution of Power Spectral Estimates for Noise and Clutter Data

Experiments consisting of examining the statistical distributions of the estimated power spectral density at various Doppler frequencies for different realizations of input noise data (white or colored) were made. It was found in general that, for white gaussian noise, regardless of the order of the PEF derived or the number of data analyzed, the estimated values of the spectral density at any given frequency were log-normally distributed; i.e., the logarithms of the spectral density estimates at any given Doppler frequency had a gaussian or normal distribution. Such a distribution has a long "tail", so that if use of the Burg algorithm as a detector is comtemplated, then detection thresholds must be set very high in order to achieve low probability of false alarm (P_{FA}). The fact that such a probability distribution function describes the processor output statistics also implies that large dynamic range capability within the processor digital hardware may be required.

For white gaussian noise inputs, it was found that the mean power spectral density level increased as N, the number of data, was increased, and decreased as M, the order of the derived PEF, was increased. Also it was found that the variance of the power spectral density decreased as N was increased, and increased as M was increased.

Reference to eqns. (3.23) or (3.35), and (3.41) suggests these results are intuitively reasonable. If successive data samples are uncorrelated, and have zero mean, then the expected value for $\beta(K,K)$ is zero, and $\langle|\beta(K,K)|^2\rangle$ and $\langle|\beta(K,K)|^4\rangle$ ought to vary roughly as N^{-2} and N^{-4} respectively. Examination of eqns. (3.36) and (3.37) shows that the $f_K(n)$s and the $b_K(n)$s

ought also to be gaussian white noise. Then consideration of eqns. (3.39) and (3.41) suggests that, since the expected values of the PEF coefficients are zero for white noise input, the fluctuations in the mean power level and the variance of spectra are due principally to fluctuations in the means and variances of the Π_Ks, which in turn are related directly to the means and variances of the $|\hat{\beta}(K,K)|^2$s. This discussion therefore implies that, as usual, better reliability is achieved as the number of data processed is increased.

For colored noise inputs, the statistical properties of the Burg algorithm became more highly variable. It was found that at frequencies where the slope of the noise spectrum was significant, the tails of the observed statistical distributions of the estimated power spectral densities were consistently longer than those of the corresponding log-normal distributions which had the same means and variances as the observed distributions. This again suggests the necessity of very high threshold settings for acceptable P_{FA}, and the requirement for large dynamic range in the implementation of the reflection-coefficient estimator. Only in the case of observations made at a Doppler frequency where the slope of the colored noise spectrum was zero did a log-normal distribution adequately describe the observed statistics of the output power spectral densities.

5.2.2.2 Doppler Detection Characteristics of the Burg Algorithm

The characteristics of the Burg algorithm as a Doppler target detector for detecting non-fluctuating targets in the presence of interference in the form of low-level white noise (corresponding to receiver noise) plus low-pass colored noise or a DC line (corresponding to clutter) were also investigated by means of simulation studies in Ref. [4]. The performance of the Burg algorithm was compared to that of 8-point and 16-point DFTs preceded by a two-pulse canceller. The detection threshold was set for a P_{FA} of 10^{-6} (as determined by extrapolation of the no-target simulation results), and a probability of detection P_D of 0.9 was chosen.

It was found that, for the case of the DC line-clutter model only, the performance of the Burg algorithm was superior to that of the canceller-DFT processor operating on the same data for Doppler frequencies in the range of about ±0.125 of the radar PRF.

It was suggested [4] on this basis that perhaps the Burg algorithm could provide some benefit for improved Doppler detection in the region of very narrow-bandwidth clutter, but the validity of the conjecture remains to be demonstrated for clutter with small but finite bandwidths.

5.3 INVERSE CLUTTER-CORRELATION MATRIX ESTIMATION AND ADAPTIVE FILTERING

An interesting potential application for the Burg algorithm has recently been proposed in Ref. [23]. It is suggested there that the Burg algorithm be used to estimate a set of reflection coefficients based on data containing clutter returns only, and that these reflection coefficients then be used to compute an estimate of the inverse of the extrapolated clutter correlation matrix. This inverse matrix is used to generate a set of adaptive weights $\underline{w}$ for use in a maximum-likelihood Doppler filter bank, according to the well-known result

$$\underline{w} = [\hat{R}]^{-1}\,\underline{u} \tag{5.2}$$

where here $[\hat{R}]^{-1}$ is the estimated inverse clutter-correlation matrix and $\underline{u}$ is a frequency-steering vector

$$\underline{u}^{+} = \Big\{\exp[-j\pi(N-1)f\Delta t], \exp[-j\pi(N-3)f\Delta t], \ldots, \exp[j\pi(N-1)f\Delta t]\Big\} \tag{5.3}$$

As before, the superscript + denotes conjugate transposition, and N is the length of the data set and the dimension of the inverse extrapolated matrix.

A technique is also described for updating the estimated inverse matrix. The reflection coefficient estimates are updated on a data-set by data-set basis, using decaying-memory digital integration, and periodically a new estimate of the inverse matrix is made. (This averaging technique differs from that described in Section 5.2.1: there the numerator and denominator terms of eqn. (3.35) are averaged before the quotient is computed; here it is the quotients that are averaged.)

The results obtained using simulated clutter data appear encouraging. However, it remains to demonstrate the effectiveness of the technique using actual radar data. Also, schemes for avoiding the inadvertent suppression of wanted targets must be considered.

5.4 ANGULAR SPECTRUM ESTIMATION

It is well known that the angle of arrival of signals having a known carrier frequency and incident on a linear antenna array can be described by an angular spectrum analogous to the frequency spectrum describing time-domain signals. However, caution must be exercised in pursuing this analogy, for the characteristics of the time and spatially sampled signals can be quite different. For example, multiple "glints" from a moving radar target (e.g., an aircraft) as observed across an antenna aperture correspond at any instant in time to a set of very narrow angular spectral lines with very small angular separation, whereas the Doppler spectrum from such a glinting target will be much more complicated, due to quasi-random amplitude fading imposed on the echoes by the continually varying angular orientation of the moving target. Furthermore, there exists the potential for short-term time averaging in the analysis of angular spectra, the analog of which (spatial diversity) is often not available for the analysis of Doppler spectra.

Unlike the case of sonar and seismic data processing, relatively little work has been published concerning the angular spectral analysis of radar-like signals using maximum-entropy techniques. In fact, only three significant and relevant papers [24,25,26] on this topic have come to the attention of this reviewer.

King [24] has performed an interesting set of simulation studies which show both the beneficial effects of averaging angular spectra and also the occurrence of "line splitting", the troublesome phenomenon which can occur

when the Burg algorithm is applied to signals containing multiple sinusoidal signals at high SNR [27].

On the experimental side, an attempt has been made to apply the Burg algorithm to the measurement of the elevation angles of arrival of multi-component H.F. (3-30 MHz) skywaves [25] observed using a vertical antenna array comprised of eight horizontal loop elements. HF skywave signals are almost invariably afflicted with ionospherically-caused multipath components, and are additionally complicated by reflections from a non-uniform ground-plane in the near field of the antenna. While the Burg algorithm was successfully applied to simulated data modelled after a theoretical description of the expected data, the experimental data proved too much for it, although more conventional Fourier processing appeared to cope successfully.

However, there exists one report [25] of experimentally successful angular spectrum estimation at 1.257 GHz using a 12-element vertical sampled-aperture array to receive CW signals from an airborne beacon at a range of 0.4 mi (0.6 km). Clearly the SNR of such a signal would be high. Also, the direct and ground-reflected signals were separated in angle by more than the classical beamwidth of the array.

The fourth-order (M=4) Burg spectral estimate was compared to the classical Fourier transform angular spectrum and to a maximum-likelihood spectral estimate. The Burg spectrum was the easiest to interpret, consisting simply of two spikes corresponding to the direct and reflected signals. However, although it was not pointed out in the report, the peak of the Fourier transform spectrum corresponded more closely to the geometrically expected angle of arrival than did the Burg estimate. Perhaps this was another manifestation of the line-splitting and closely related line-pulling phenomenon which is sometimes encountered in high SNR situations.

6. SUMMARY AND PROGNOISIS

This report has had two main purposes. The first was to present the foundations and mathematical development of the maximum-entropy (Section 2) and Burg (Section 3) algorithms for autoregressive spectral analysis in a clear and relatively complete manner. Emphasis was on clarity of presentation, in order to remove the apparent shroud of mystery which often seems to surround these topics. For completeness, Fourier transform techniques were briefly discussed in Section 1.3 and other techniques were discussed in Section 1.4 and Section 4.

The report's second purpose was to describe and discuss areas of potential application of the Burg algorithm to radar signal processing. It appears that there are two areas of promise: clutter analysis, classification and/or suppression; and angular spectrum (angle-of-arrival) estimation. Preliminary results concerning the direct application of the Burg spectral estimate (eqn. (3.39)) as a target detector are pessimistic: it appears that a white noise input signal with gaussian statistics gives rise to outputs with log-normal statistics in the spectral domain.

6.1 CLUTTER PROCESSING

It appears for example, that if sufficient data are available from a scanning surveillance radar, e.g., 10 scans of a resolution cell with 8-20 hits per scan, then the modified [20] Burg algorithm (see Section 5.2.1) can be used to obtain a single prediction-error filter (PEF) which out to reliably represent the clutter spectrum from the particular cell under observation. The PEF can then be used to form a spectral estimate which characterizes, and thus can be used to classify, the observed clutter (e.g., ground, weather, birds). Also, the PEF can be used in a conventional digital transversal or finite-duration impulse-response (FIR) MTI whitening filter which would be updated on a regular basis and thus be adapted to suppress the observed clutter spectrum, or to adaptively design a maximum-likelihood Doppler filter bank [23] for the optimum detection of moving targets.

However, the computational complexity of the algorithm, as discussed in Section 5.1 et seq., indicates that clutter estimation could likely be done on selected cells only, so that some sort of area clutter mapping procedure would be required to decide what cells to process, and over what adjacent coverage areas the results could be validly extrapolated. This, the avoidance of suppressing wanted targets, and the selection of the appropriate order of the PEF are but three aspects of this area of application which require further, experimental study.

6.2 ANGULAR SPECTRUM (ANGLE-OF-ARRIVAL) ANALYSIS

Although the Burg algorithm as described here is applicable only to single-channel equi-spaced sampled data as would be obtained from a linear sampled-aperture array, it appears that the application of the algorithm to such data warrants further investigation. In particular, the option of appropriately time-averaging individual "frames" of array data may mean that good statistical reliability may be obtainable in relatively short time intervals. However, it is not clear at present whether or in what context sufficient benefits could be gained to warrant the complexity of processing required to implement the algorithm.

7. ACKNOWLEDGEMENT

This work is supported by the Department of National Defence under Research and Development Branch Project No. 33C69.

8. REFERENCES

1. Blackman, R.B. and T.W. Tukey, *The Measurement of Power Spectra*, Dover Publications, Inc., New York, 1959.

2. Baggeroer, A.B., *Confidence Intervals for Regression (MEM) Spectral Estimates*, IEEE Trans. Inform. Theory, IT-22, 534-545, September 1976.

3. Kesler, S.B., *Nonlinear Spectral Analysis of Radar Clutter*, Report CRL-51, Communications Research Laboratory, Faculty of Engineering, McMaster University, Hamilton, Canada, January 1978.

4. Chan, H.C. and S. Haykin, *Applications of the Maximum Entropy Method in Radar Signal Processing*, Report CRL-62, Communications Research Laboratory, Faculty of Engineering, McMaster University, Hamilton, Canada, March 1979.

5. Haykin, S., S. Kesler and B. Currie, *An Experimental Calssification of Radar Clutter*, Proc. IEEE, 67, 332-333, February 1979.

6. Burg, J.P., *Maximum Entropy Spectral Analysis*, presented to the 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, OK., 31 October 1967.

7. Burg, J.P., *A New Analysis Technique for Time Series Data*, presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands, 12-23 August 1968.

8. Burg, J.P., *Maximum Entropy Spectral Analysis*, Ph.D. Thesis, Stanford University, Stanford, CA., May 1975.

9. Shannon, C.E. and W. Weaver, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, IL, 1959.

10. Ulrych, T.J. and T.N. Bishop, *Maximum Entropy Spectral Analysis and Autoregressive Decomposition*, Rev. Geophysics and Space Phys., 13, 183-200, February 1975.

11. Nuttal, A.H., *Spectral Analysis of a Univariate Process with Bad Data Points, Via Maximum Entropy and Linear Prediction Techniques*, NUSC Technical Report 5303, Naval Underwater Systems Center, New London Laboratory, 26 March 1976, AD-A024942.

12. Akaike, H., *Autoregressive Model Fitting for Control*, Ann. Inst. Statist. Math., 23, 163-180, 1971.

13. Akaike, H., *A New Look at the Statistical Model Identification*, IEEE Trans. Autom. Control, AC-19, 716-723, December 1974.

14. Landers, T.E. and R.T. Lacoss, *Some Geophysical Applications of Autoregressive Spectral Estimates*, IEEE Trans. Geosci. Electron, GE-15, 26-32, January 1977.

15. Parzen, E., *An Approach to Time Series Modelling and Forecasting Illustrated by Hourly Electricity Demands*, Statist. Sci. Division, State Univ. New York, Buffalo, N.Y., Tech. Report 37, January 1976.

16. Ulrych, T.J. and R.W. Clayton, *Time Series Modelling and Maximum Entropy*, Physics of Earth and Planetary Interiors, 12, 188-200, 1976.

17. Morf, M., B. Dickinson, T. Kailath and A. Vieira, *Efficient Solution of Covariance Equations for Linear Prediction*, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-25, 429-433, October 1977.

18. Andersen, N., *Comments on the Performance of Maximum Entropy Algorithms*, Proc. IEEE, 66, 1581-1582, November 1978.

19. Kaveh, M, and G.R. Cooper, *An Empirical Investigation of the Properties of the Autoregressive Spectral Estimator*, IEEE Trans. Inform. Theory, IT-22, 313-323, May 1976.

20. Moorcroft, D.R., *Maximum-entropy Spectral Analysis of Radio-auroral Signals*, Radio Science, 13, 649-660, July-August 1978.

21. Kesler, S. and S. Haykin, *A Comparison of the Maximum Entropy Method and the Periodogram Method Applied to the Spectral Analysis of Computer-simulated Radar Clutter*, Can. Elec. Eng. J. 3, No. 1, 11-16, 1978.

22. Taylor, R.G., T.S. Durrani and C. Goutis, *Block Processing in Pulse Doppler Radar*, RADAR-77, IEE Conference Publication No. 155, 373-378, 25-28 October 1977.

23. Sawyers, J.H., *Applying the Maximum Entropy Method to Adaptive Digital Filtering*, Conference Record: Twelth Asilomar Conference on Circuits, Systems and Computers, 6, 7, 8 November 1978, (The author is with Hughes Aircraft Company, Fullerton, California.)

24. King, W.R., *Antenna Patterns Computed with Maximum Entropy and the Burg Technique*, Proceedings of the RADC Spectrum Estimation Workshop, 24, 25 and 26 May 1978, AD-A054650.

25. Clarke, R.H. and D.V. Tibble, *Measurement of the Elevation Angles of Arrivals of Multicomponent H.F. Skywaves*, Proc. IEE, 125, 17-24, January 1978.

26. Evans, J.E., *Aperture Sampling Techniques for Precision Direction Finding*, Electro/78 Conference Record, Presented at Electro/78, Boston, MA, 23-25 May 1978, Paper 28/3 (8 pp).

27. Herring, R.W., *A Comparison of the Burg and the Known-Autocorrelation Autoregressive Spectral Analysis of Complex Sinusoidal Signals in Additive White Noise*, CRC Report 1326, October 1979.

NOTE: References [2,6,7,10,13,14,16,17 and 19] have been reprinted in *Modern Spectrum Analysis*, Edited by Donald G. Childers, IEEE Press, New York, 1978.